

به نام خدا



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



درس هوش مصنوعی قابل اعتماد
مدرس: دکتر مصطفی توسلی‌پور

تمرین شماره ۳

اردیبهشت ماه ۱۴۰۳

۳ مقدمه
۳ سوال اول:
۴ زیربخش اول
۴ زیربخش دوم
۴ سوال دوم:
۵ سوال سوم:
۶ زیربخش اول
۶ زیربخش دوم
۶ زیربخش سوم
۶ زیربخش چهارم
۷ زیربخش پنجم
۷ سوال چهارم:
۸ سوال پنجم:
۸ زیربخش اول
۸ زیربخش دوم
۸ زیربخش سوم
۸ زیربخش چهارم
۹ زیربخش پنجم ()
۹ زیربخش ششم
۹ سوال ششم:
۹ زیربخش اول
۹ زیربخش دوم
۱۰ نکات تحویل

مقدمه

در این تمرین، هدف اصلی درک بهتر روابط علی میان متغیرها و تفسیر کردن مدل‌ها با استفاده از روتبط علس بین متغیرهاست. به عبارت دیگر، با بکارگیری علّیت در تفسیرپذیری، می‌توانیم درک عمیق‌تری از پدیده‌های پیچیده کسب کنیم و تصمیمات مناسب‌تری برای مداخلات احتمالی اتخاذ نماییم.

سوال اول

فرض کنید $S \in \{L, R\}$ اندازه سنگ کلیه بیمار را نشان می‌دهد، $S = L$ نشان‌دهنده‌ی بزرگ بودن سنگ کلیه و R بیانگر معمولی بودن آن است. همچنین $A \in \{N, O\}$ نشان‌دهنده وضعیت بیمار است که در آن $A = N$ نشان‌دهنده این است که بیمار درمان قدیمی را دریافت می‌کند. در نهایت، $Y \in \{0, 1\}$ نشان می‌دهد که حذف یا برجا بودن سنگ کلیه بعد از یک ماه است بدین صورت که $Y = 1$ نشان‌دهنده‌ی حذف سنگ کلیه و $Y = 0$ نشان می‌دهد برجا بودن آن است.

با فرض تجزیه $P_X(S, A, Y)$ با توجه به DAG ارائه شده در شکل 1 و اینکه مقادیر مربوط به توزیع‌های مشروط به صورت زیر می‌باشند،

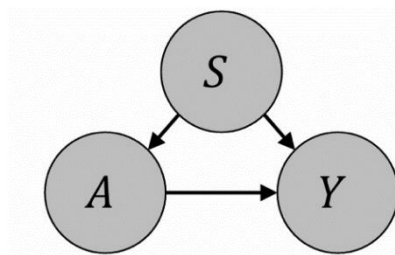


Figure 1 DAG

$$P_X(S = L) = 0.49$$

$$P_X(A = N \mid S = L) = 0.77$$

$$P_X(A = N \mid S = R) = 0.24$$

$$P_X(Y = 1 \mid S = L, A = N) = 0.73$$

$$P_X(Y = 1 \mid S = L, A = O) = 0.69$$

$$P_X(Y = 1 \mid S = R, A = N) = 0.93$$

$$P_X(Y = 1 \mid S = R, A = O) = 0.87$$

زیربخش اول (5 نمره)

مقادیر $P_X(Y = 1 | A = N)$ و $P_X(Y = 1 | A = O)$ را محاسبه کنید.

زیربخش دوم (5 نمره)

مقادیر $P_X(Y = 1 | do(A = N))$ و $P_X(Y = 1 | do(A = O))$ را محاسبه کنید.

سوال دوم

تقاضای درخواست وام دو فرد A, B توسط بانک رد می‌شود. با در نظر گرفتن روابط مربوط به SCM و توزیع‌های مربوط به تولید داده که در ادامه آمده‌است، برای classifier ارائه شده، با استفاده از مسئله Causal Algorithmic recourse، بهینه‌ترین حالت ممکن را در آن وام به آن‌ها تعلق می‌گرفت و هزینه مربوط به تغییر هر کدام را بیابید. ویژگی‌های مربوط به این دو فرد به صورت مقابل می‌باشد. (12 نمره)

$$A = [\$75000, \$25000]^T$$

$$B = [\$70000, \$23800]^T$$

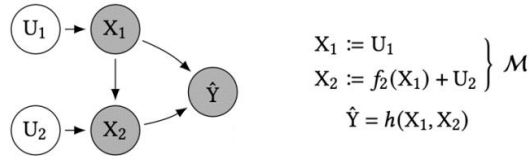


Figure 2 Graph and SCMs in which X_1 , X_2 and Y represent annual salary, bank balance and the classifier outcome indicating the eligibility to receive loan.

$$U_1 \sim \$10,000 \cdot \text{Poisson}(10),$$

$$U_2 \sim \$2,500 \cdot \mathcal{N}(0, 1),$$

$$X_2 := \frac{3}{10} \cdot X_1 + U_2,$$

$$h = \text{sgn}(X_1 + 5 \cdot X_2 - \$225,000)$$

سوال سوم

داده‌هایی که در اختیارتان قرار گرفته است مربوط به سیستم مدیریت عملیات برای یک آژانس هواپیمایی در طی یک سال می‌باشد. با استفاده از این داده‌ها، مدیران آژانس هواپیمایی می‌توانند عملکرد روزانه خود را در زمینه فروش بلیط، جذب مشتری، کنترل هزینه‌ها و سودآوری، ارزیابی کنند. آن‌ها می‌توانند تصمیمات مناسب را برای بهبود استراتژی‌های بازاریابی، ارتقای تجربه مشتری، مدیریت هزینه‌ها و در نهایت افزایش سودآوری اتخاذ نمایند.

هر یک از ویژگی‌های ثبت شده برای این داده‌ها به شرح زیر است:

- **Booking_Mode**: نشان می‌دهد که آیا آن روز رزرو بلیط هواپیما به علت خاصی رخ داده است یا خیر مثلاً آن روز تعطیلات ویژه‌ای در تقویم هست و غیره، این ویژگی بر روی ویژگی‌های زیر اثر می‌گذارد:
 - Marketing_Budget ○
 - Website_Visits ○
 - Tickets_Sold ○
 - Ticket_Price ○
- **Marketing_Budget**: میزان هزینه‌های صرف شده برای تبلیغات و بازاریابی خدمات آژانس هواپیمایی در آن روز، مانند تبلیغات در رسانه‌ها، شبکه‌های اجتماعی و غیره، این ویژگی بر روی ویژگی‌های زیر اثر می‌گذارد:
 - Website_Visits ○
 - Operating_Expenses ○
- **Website_Visits**: تعداد بازدیدکنندگان وبسایت آژانس هواپیمایی در آن روز، این ویژگی بر روی ویژگی‌های زیر اثر می‌گذارد:
 - Tickets_Sold ○
- **Ticket_Price**: قیمت متوسط هر بلیط هواپیما فروخته شده توسط آژانس در آن روز، این ویژگی بر روی ویژگی‌های زیر اثر می‌گذارد:
 - Tickets_Sold ○
 - Sales_Revenue ○
- **Tickets_Sold**: تعداد کل بلیط‌های هواپیما فروخته شده توسط آژانس در آن روز، این ویژگی بر روی ویژگی‌های زیر اثر می‌گذارد:
 - Sales_Revenue ○
 - Operating_Expenses ○
- **Sales_Revenue**: درآمد کسب شده از فروش بلیط‌های هواپیما در آن روز، این ویژگی بر روی ویژگی‌های زیر اثر می‌گذارد:
 - profit ○

- **Operating_Expenses**: شامل تمامی هزینه‌های عملیاتی آژانس هواپیمایی از جمله هزینه‌های اداری، هزینه‌های نگهداری سیستم‌ها و نرم‌افزارها، حقوق کارکنان، اجاره دفتر و غیره در آن روز، این ویژگی بر روی ویژگی‌های زیر اثر می‌گذارد:
 - profit

- **Profit**: سود این آژانس هواپیمایی پس از کسر هزینه‌های عملیاتی از درآمد کسب شده در آن روز.

با استفاده از دیتاست داده شده مراحل زیر را طی کرده و به سوالات پاسخ دهید. (توجه شود بخش عمده‌ای از نمره شما از این سوال به گزارش و تحلیل آن وابسته است)

زیربخش اول (۲ نمره)

با توجه به ویژگی‌های داده شده و با استفاده از کتابخانه `networkx`، گراف علی را رسم کنید.

زیربخش دوم (۵ نمره)

یک مدل علی ساختاری^۱ را براساس گراف علی رسم شده در قسمت (الف) مدل‌سازی کنید. در این مدل، هر متغیر را به صورت یک تابع از متغیرهای علی (والدین) خود، به اضافه یک جمله نویز یا خطا در نظر بگیرید. برای مدل‌سازی توابع تولیدکننده هر متغیر از والدین آن، می‌توانید از رگرسیون خطی، درخت‌های تصمیم یا هر مدل دیگری که مناسب تشخیص می‌دهید، استفاده کنید. توزیع احتمالی جمله نویز یا خطا را نیز باید مشخص کنید (می‌توانید از توزیع نرمال یا هر توزیع دیگری که برای داده‌هایتان مناسب است، بهره بگیرید). سپس با استفاده از داده‌های موجود، پارامترهای مدل را برازش کنید.

پس از مدل‌سازی و برازش مدل علی ساختاری، به سوالات علی مطرح شده در ادامه سوال پاسخ دهید.

زیربخش سوم (۳ نمره)

واریانس سود را به دست بیاورید، سپس با توجه به مقدار آن و با نگاهی به نمودار علی، مشاهده می‌شود که درآمد و هزینه عملیاتی، تأثیر مستقیمی بر سود دارند، اما کدام یک از آن‌ها بیشترین تأثیر را در واریانس سود دارند؟ (**سهم هر یک از دو متغیر گفته شده را در سود را بیان کنید)

آیا نتایج به دست آمده منطقی است؟ چرا؟

زیربخش چهارم (۵ نمره)

در قسمت قبل بررسی کردیم واریانس سود با چه نسبتی، از هر یک از والدین آن ناشی می‌شود، در این بخش بررسی کنید، کدام یک از عوامل سیستم بیشترین تأثیر را در ایجاد این میزان از واریانس سود دارند.

¹ structural causal model

زیربخش پنجم (۵ نمره)

در ابتدای اولین روز سال جدید مقادیر زیر مشاهده شده است:

جدول ۱ مقادیر مشاهده شده در اولین روز سال جدید

Bookin g_mod e	Marketin g_Budg et	Websit e_Visit s	Ticke t_Pric e	Ticket s_Sol d	Sales_ Reven ue	Operating _Expense s	Profit
True	\$2,079.0 1	21110	\$700. 47	7987	\$5,594, 652.87	\$4,495,58 8.74	\$1,099 ,064.1 3

مقدار سود در این روز نسبت به روز اول سال گذشته کاهش داشته است یا افزایش؟ عوامل کاهش یا افزایش سود در این روز چیست؟ استدلال خود را در رابطه با این نتیجه بیان کنید.

سوال چهارم

در این سوال قصد داریم که اثر insulin بر blood glucose را تخمین بزنیم. فرض کنید رابطه‌ی بین Age و blood glucose and blood pressure، insulin به صورت شکل زیر می باشد.

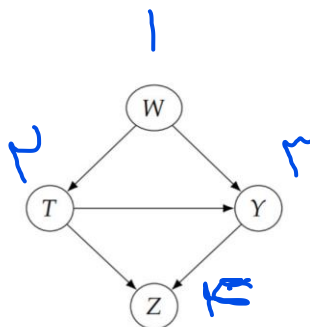


Figure ۳ Causal relationships between W, T, Y, Z variables demonstrating Age, insulin level, blood glucose and blood pressure, respectively.

با استفاده از مقادیر مربوط به داده های ارائه شده در فایل 'health.csv' و با استفاده از logistic regression مقادیر زیر را محاسبه کنید. (12 نمره)

$$E_{W,Z} E_Y[y | t, W, Z]$$

$$E_W E_Y[y | t, W]$$

$$E_Y[y | t]$$

کدام یک از موارد بالا که اثر علی insulin بر blood glucose را نشان می دهد؟ ثابت کنید. (10 نمره)

سوال پنجم:

در این سوال قصد داریم با استفاده از داده‌های داده‌شده در سوال قبل، ابتدا با طراحی یک طبقه‌بند، دو گروه سالم و ناسالم را از هم جدا نموده و سپس با استفاده از دو روش Nearest Counterfactual explanation و Casual Algorithmic recourse، بهترین و به‌صرفه‌ترین مداخله برای تغییر وضعیت افراد بیمار به سالم را بیابیم. (20 نمره)

بدین منظور، گراف مربوط به سوال 4 و داده‌های 'health.csv' را در نظر بگیرید. در این فایل داده‌های مربوط به Age، insulin، blood glucose، blood pressure، بیانگر ورودی‌های طبقه‌بند و ستون category بیانگر سالم یا ناسالم بودن فرد مورد نظر در خروجی classifier است (0 بیانگر ناسالم بودن و 1 بیانگر سالم بودن فرد مورد نظر است).

با انجام مراحل زیر، پس از طراحی classifier به مقایسه‌ی دو روش مربوطه می پردازیم.

زیربخش اول

به فایل 'data-utils.py' را مراجعه نموده و تابع process_health_data را طوری تکمیل نمایید که:

۱. insulin، blood glucose تنها ویژگی‌های actionable باشند.
۲. مقادیر blood pressure، insulin، blood glucose از حداکثر و حداقل مقادیر موجود در دیتاست فراتر نروند.

زیربخش دوم

'main.py' را به ازای 10 فرد ناسالم اجرا کنید و هزینه محاسبه شده را گزارش کنید. این هزینه، چه چیزی را نشان می‌دهد؟

زیربخش سوم

با مراجعه به فایل 'scm.py'، کلاس Health_SCM را کامل کنید به طوری که ویژگی‌های insulin، blood glucose تنها ویژگی‌های actionable، blood pressure و Age ویژگی‌های constant باشند.

زیربخش چهارم

با توجه به ضرایب SCM، تابع get_Jacobian را به گونه ای تکمیل کنید که ژاکوبین SCM را به عنوان خروجی برگرداند.

زیربخش پنجم

Comment های مربوط به کدهای comment شده در تابع get_scm را در 'utils.py' حذف کنید. دوباره 'main.py' را برای 10 فرد ناسالم اجرا کنید و هزینه محاسبه شده را گزارش کنید. هزینه محاسبه شده، چه چیزی را نشان می دهد؟

زیربخش ششم

هزینه های محاسبه شده از قسمت B و E را مقایسه کنید. کدام یک کمتر است؟ چرا؟ یکی از افراد ناسالم را در نظر بگیرید و هزینه action های مربوطه، و متغیرهایی که بر روی آنها مداخله شده را برای هر دو قسمت دوم و پنجم را گزارش نموده و مقایسه کنید.

سوال ششم:

مقاله 'On the Adversarial Robustness of Causal Algorithmic Recourse' را بررسی کنید و به سوالات زیر پاسخ دهید.

زیربخش اول (8 نمره)

تحت چه شرایطی در طبقه بندی کننده و SCM، Robustness در Causal Algorithmic Recourse تضمین می شود؟ به طور شهودی دلایل را توضیح دهید.

زیربخش دوم (8 نمره)

با توجه به نتایج مربوط به Proposition 4 در مقاله، به طور شهودی دلایل استخراج معادله 5 مقاله را توضیح دهید.

نکات تحویل

- مهلت ارسال این تمرین تا پایان روز "۱۰ خرداد ماه" خواهد بود.
 - این زمان قابل تمدید نیست و در صورت نیاز می‌توانید از grace time استفاده کنید.
 - در نظر داشته باشید که حداکثر مهلت آپلود تمرین در سامانه تا ۷ روز پس مهلت تحویل است و پس از آن سامانه بسته خواهد شد.
 - پیاده سازی با زبان برنامه نویسی پایتون باید باشد و کدهای شما باید قابل اجرا بوده و به همراه گزارش آپلود شوند.
 - انجام این تمرین به صورت یک نفره می‌باشد.
 - در صورت مشاهده هر گونه تشابه در گزارش کار یا کدهای پیاده‌سازی، این امر به منزله تقلب برای طرفین در نظر گرفته خواهد شد.
 - استفاده از کدهای آماده بدون ذکر منبع و بدون تغییر به منزله تقلب خواهد بود و نمره تمرین شما صفر در نظر گرفته می‌شود.
 - در صورت رعایت نکردن فرمت گزارش کار نمره گزارش به شما تعلق نخواهد گرفت.
 - تحویل تمرین به صورت دستنویس قابل پذیرش نیست.
 - تمامی تصاویر و جداول مورد استفاده در گزارش کار باید دارای توضیح (caption) و شماره باشند.
 - بخش زیادی از نمره شما مربوط به گزارش کار و روند حل مسئله است.
 - لطفا گزارش، فایل کدها و سایر ضmann مورد نیاز را با فرمت زیر در سامانه بارگذاری نمائید.
 - HW1_[Lastname]_[StudentNumber].zip
 - در صورت وجود سوال و یا ابهام می‌توانید از طریق رایانامه زیر با موضوع TAI_HW1 با دستیاران آموزشی در ارتباط باشید:
 - سوال اول، سوم
- Fatemehnnadi@ut.ac.ir
- سوال دوم، چهارم، پنجم و ششم
- mhdhshri@gmail.com
- با آرزوی سلامتی و موفقیت روزافزون.